(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 2 575 139 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
03.04.2013 Bulletin 2013/14

(51) Int Cl.:
G11C 16/04 *(2006.01)* ***H01L 29/792*** *(2006.01)*

(21) Application number: **12168805.5**

(22) Date of filing: **22.05.2012**

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME

(30) Priority: **30.09.2011 US 201161541085 P**

(71) Applicant: **eMemory Technology Inc.**
Hsin-Chu
30075 (TW)

(72) Inventors:
- **Hsiao, Kai-Yuan**
  408 Taichung City (TW)
- **Lee, Wen-Yuan**
  302 Hsinchu County (TW)
- **Ting, Yun-Jen**
  302 Hsinchu County (TW)
- **Liu, Cheng-Jye**
  302 Hsinchu County (TW)
- **Sun, Wein-Town**
  325 Taoyuan County (TW)

(74) Representative: **Becker Kurig Straus**
Patentanwälte
Bavariastrasse 7
80336 München (DE)

(54) Method of programming nonvolatile memory

(57) Each memory cell (10) of a plurality of memory cells (130A-130I) of a memory has a well, source (132) and drain (133) regions, a storage layer (135), and a gate (131). The memory cells (130A-130I) are in a matrix. Same column drain regions (133) connect to the same bit line (110A-110C), same row gates (131) connect to the same word line (100A-100C), and same column source regions (132) connect to the same source line (120A-120C). The memory (10) is programmed by applying a first voltage to a word line (100A-100C) electrically connected to a memory cell (130E) of the plurality of memory cells (10), applying a second voltage different from the first voltage by at least a programming threshold to a bit line (110A-110C) electrically connected to the memory cell (130E), applying a third voltage different from the first voltage by at least the programming threshold to a source line (120A-120C) electrically connected to the memory cell (130E), and applying a substrate voltage to the plurality of memory cells (10, 130E).

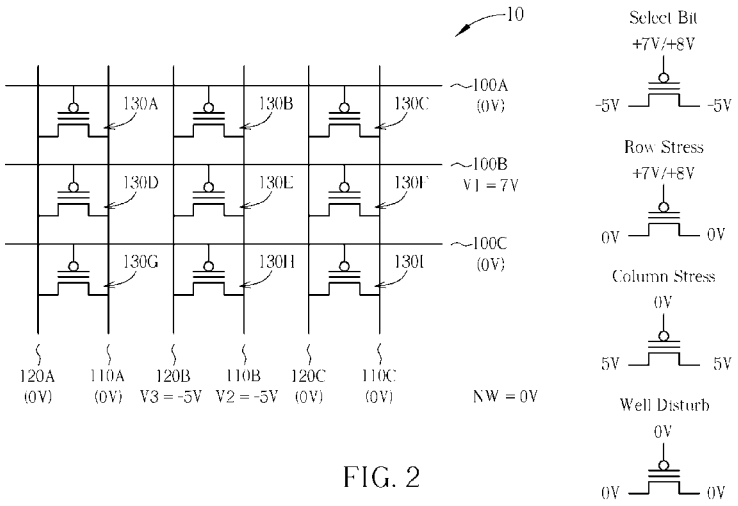

FIG. 2

EP 2 575 139 A2

Description

Field of the Invention

[0001] The present invention relates to nonvolatile memory, and more particularly to nonvolatile memory and a method of programming the nonvolatile memory according to the pre-characterizing clause of claims 1 and 8.

Background of the Invention

[0002] In current nonvolatile memory applications, critical dimension sensitivity is a barrier to improved performance. One programming method that exhibits lower sensitivity to critical dimension is Fowler-Nordheim (FN) programming, which has good uniformity, and thus lower critical dimension sensitivity. However, FN programming is also slow. Hot carrier programming, such as band-to-band hot electron (BBHE) programming is faster, but is more sensitive to critical dimension, which limits its application as a nonvolatile memory programming technique.

Summary of the Invention

[0003] This in mind, the present invention aims at providing a nonvolatile memory that can be programmed quickly while maintaining low critical dimension sensitivity.

[0004] This is achieved by a nonvolatile memory according to claims 1 and 8. The dependent claims pertain to corresponding further developments and improvements.

[0005] As will be seen more clearly from the detailed description following below, the claimed nonvolatile memory includes a novel architecture that has a gate terminal connected to a word line, a drain terminal connected to a bit line, and a source terminal connected to a source line where the source lines are not electrically connected to each other.

[0006] According to an embodiment, a programming method is for use in memory. Each memory cell of a plurality of memory cells of the memory comprises a well having a first conductivity type, source and drain regions of a second conductivity type opposite the first conductivity type, a storage layer, and a gate. The plurality of memory cells are arranged in a matrix with drain regions in the same column electrically connected to the same bit line, gates in the same row electrically connected to the same word line, and source regions in the same column connected to the same source line. The method comprises applying a first voltage to a word line electrically connected to a memory cell of the plurality of memory cells, applying a second voltage different from the first voltage by at least a programming threshold to a bit line electrically connected to the memory cell, applying a third voltage different from the first voltage by at least the programming threshold to a source line electrically connected to the memory cell, and applying a substrate voltage to the plurality of memory cells.

[0007] According to an embodiment, a memory array comprises a bit line electrically connected to a column of memory cells of the memory array, a source line electrically connected to the column of memory cells of the memory array, and not electrically connected to other source lines of the memory array, a word line electrically connected to a row of memory cells of the memory array, and a memory cell. The memory cell comprises a gate terminal electrically connected to the word line for receiving a first voltage during a programming operation, a first terminal electrically connected to the bit line for receiving a second voltage different from the first voltage by at least a programming threshold during the programming operation, a second terminal electrically connected to the source line for receiving a third voltage different from the first voltage by at least the programming threshold during the programming operation, and a well terminal for receiving a substrate voltage during the programming operation.

Brief Description of the Drawings

[0008] In the following, the invention is further illustrated by way of example, taking reference to the accompanying drawings thereof:

FIG. 1 is a diagram of a memory array according to an embodiment;
FIG. 2 is a diagram illustrating programming of memory cell 130E of memory array;
FIG. 3 is a diagram showing voltages of a read operation on memory cell of memory array;
FIG. 4 is a diagram showing layout of the memory array of FIG. 1; and
FIG. 5 is a diagram showing layout of the memory array 10 of FIG. 1.

Detailed Description

[0009] Please refer to FIG. 1, which is a diagram of a memory array 10 according to an embodiment. The memory array 10 comprises a plurality of word lines 100A-100C, a plurality of bit lines 110A-110C, a plurality of source lines 120A-120C, and a plurality of memory cells 130A-130I. Only nine memory cells are shown in FIG. 1 for simplicity of description. In practice, the memory array 10 may comprise many more memory cells than the nine shown in FIG. 1.

[0010] Taking memory cell 130E as an example, memory cell 130E may be a metal-oxide-semiconductor (MOS) transistor, such as a P-type MOS (PMOS), or an N-type MOS (NMOS) transistor, and may include a stack dielectric storage layer. Memory cell 130E has a gate terminal electrically connected to word line 100B, a drain terminal electrically connected to bit line 110B, and a

source terminal electrically connected to source line 120B. Source lines 120A-120C are not electrically connected to each other.

[0011] Please refer to FIG. 2 and FIG. 3. FIG. 2 is a diagram illustrating programming of memory cell 130E of memory array 10. FIG. 3 is a diagram of memory cell 130E during programming. Memory cell 130E comprises a polysilicon gate 131, a source region 132, a drain region 133, a substrate region 134, and a charge trapping layer 135. To program memory cell 130E, in the case of the memory cell 130E being a PMOS transistor, a high voltage V1 is applied to the gate 131 of the memory cell 130E through the word line 100B, a first low voltage V2 is applied to the drain 133 of the memory cell 130E through the bit line 110B, and a second low voltage V3 is applied to the source 132 of the memory cell 130E through the source line 120B. For example, the high voltage V1 may be 7 Volts, the first low voltage V2 may be -5 Volts, and the second low voltage V3 may be -5 Volts. In this example, the first low voltage V2 and the second low voltage V3 are the same. The first low voltage V2 and the second low voltage V3 may also be different. For example, a programming threshold may be set that indicates voltage difference sufficient for programming the memory cell 130E, e.g. through BBHE programming. The programming threshold may be 7 Volts. Then, the first low voltage V2 may be set to -5 Volts, and the second low voltage V3 may be set to -3 Volts. Thus, the first low voltage V2 need only be sufficiently different from the high voltage V1 to exceed the programming threshold, and the second low voltage V3 need only be sufficiently different from the high voltage V1 to exceed the programming threshold. Using the above scheme, hot electrons 136, 137 tunnel to the charge trapping layer 135, and holes 138, 139 are drawn down to the substrate 134. During programming of the memory cell 130E, all other word lines, bit lines, and source lines may be set to 0 Volts.

[0012] As shown in FIG. 2, during programming, cells of the array 10 other than the cell under programming (cell 130E in FIG. 2) are in different states. Cells 130D and 130F experience row stress, with the high voltage V1 applied to the gate terminal, and 0V applied to the source and drain terminals. Cells 130B and 130H experience column stress, with 0V applied to the gate terminal, and the first and second low voltages V2, V3 applied to the drain and source terminals, respectively. Cells 130A, 130C, 130G, and 130I experience well disturb, with all terminals (gate, drain, source) set to 0V.

[0013] Please refer to FIG. 4, which is a diagram showing voltages of a read operation on memory cell 130E of memory array 10. During reading, voltage V1 is set to 0 Volts, voltage V2 is set to -1.5 Volts, and voltage V3 is set to 0 Volts. All other source and bit line voltages are set to 0 Volts, and all other word line voltages are set to 4 Volts. Read current is read from the memory cell 130E to determine programmed state (1 or 0) of the memory cell 130E.

[0014] Please refer to FIG. 5, which is a diagram showing layout of the memory array 10 of FIG. 1. The word lines 100A-100C may be formed of polysilicon, and the bit lines 110A-110C and source lines 120A-120C may be formed in a metal layer, such as Metal 1 (M1) of a typical process. Contacts 140 electrically connect the bit lines 110A-110C to the respective drain regions of the corresponding memory cells 130A-130I, and electrically connect the source lines 120A-120C to the respective source regions of the corresponding memory cells 130A-130I. Source lines 120A-120C are not mutually electrically connected. The memory array 10 may be built in a well having a first conductivity type, e.g. N-type. The source and drain regions of each memory cell 130A-130I may be of a second conductivity type opposite the first conductivity type, e.g. P-type.

[0015] Using the above architecture and programming method, the memory array 10 has twin-side BBHE speed which is insensitive to polysilicon critical dimension (CD) variation. Program disturbance is also very slight. BBHE programming also achieves good program erase window in the memory array 10.

Claims

1. A programming method for use in memory (10), each memory cell (130A-130I) of a plurality of memory cells (130A-130I) of the memory (10) comprising a well having a first conductivity type, source (132) and drain (133) regions of a second conductivity type opposite the first conductivity type, a storage layer (135), and a gate (131), **characterized in that** the plurality of memory cells (130A-130I) being arranged in a matrix with drain regions (133) in a same column electrically connected to a same bit line (110A-110C), gates (131) in a same row electrically connected to a same word line (100A-100C), and source regions (132) in a same column connected to a same source line (120A-120C), the method comprising:

   applying a first voltage to a word line (100A-100C) electrically connected to a memory cell (130A-130I) of the plurality of memory cells (130A-130I);
   applying a second voltage different from the first voltage by at least a programming threshold to a bit line (110A-110C) electrically connected to the memory cell (130A-130I);
   applying a third voltage different from the first voltage by at least the programming threshold to a source line (120A-120C) electrically connected to the memory cell (130A-130I); and
   applying a substrate voltage to the plurality of memory cells (130A-130I).

2. The method of claim 1, **further characterized by** applying the second voltage at a same level as the third voltage.

3. The method of claim 1, **further characterized by** applying the second voltage at a different level than the third voltage.

4. The method of claim 1, wherein the first conductivity type is N-type, the second conductivity type is P-type, further **characterized by** applying the second voltage at a lower level the than the first voltage, and applying the third voltage at a lower level than the first voltage.

5. The method of claim 1, **further characterized by** inducing band-to-band hot electron (136, 137) programming of the memory cell (130A-130I) utilizing the programming threshold.

6. The method of claim 1, **further characterized in that** source lines (120A-120C) of the memory (10) are not electrically connected together.

7. The method of claim 1, **further characterized by** utilizing a stack dielectric storage layer (135) as the storage layer (135).

8. A memory array (10) comprising:

a bit line (110A-110C) electrically connected to a column of memory cells (130A-130I) of the memory array (10) ;

characterized by:

a source line (120A-120C) electrically connected to the column of memory cells (130A-130I) of the memory array (10), and not electrically connected to other source lines (110A-110C) of the memory array (10); and
a word line (100A-100C) electrically connected to a row of memory cells (130A-130I) of the memory array (10);
wherein each of the memory cells (130A-130I) comprises:

a gate terminal (131) electrically connected to the word line (100A-100C) for receiving a first voltage during a programming operation;
a first terminal (140) electrically connected to the bit line (110A-110C) for receiving a second voltage different from the first voltage by at least a programming threshold during the programming operation;
a second terminal (140) electrically connected to the source line (120A-120C) for receiving a third voltage different from the first voltage by at least the programming threshold during the programming operation; and
a well terminal for receiving a substrate voltage during the programming operation.

9. The memory array (10) of claim 8, **further characterized in that** the memory array (10) is configured such that the second voltage is the same as the third voltage.

10. The memory array (10) of claim 8, **further characterized in that** the memory array (10) is configured such that the second voltage is different from the third voltage.

11. The memory array (10) of claim 8, **further characterized in that** the well terminal is electrically connected to a well region of the memory cell (130A-130I) having a first conductivity type, the first terminal (140) is electrically connected to a source region (132) of a second conductivity type opposite the first conductivity type, and the second terminal (140) is electrically connected to a drain region (133) of the second conductivity type.

12. The memory array (10) of claim 11, **further characterized in that** the first conductivity type is N-type, the second conductivity type is P-type, the second voltage is lower than the first voltage, and the third voltage is lower than the first voltage.

13. The memory array (10) of claim 8, **further characterized in that** the programming threshold is sufficiently large to induce band-to-band hot electron (136, 137) programming of the memory cell (130A-130I).

14. The memory array (10) of claim 8, **further characterized in that** the memory cell (130A-130I) further comprises a storage layer (135).

15. The memory array (10) of claim 14, **further characterized in that** the storage layer (135) is a stack dielectric storage layer (135).

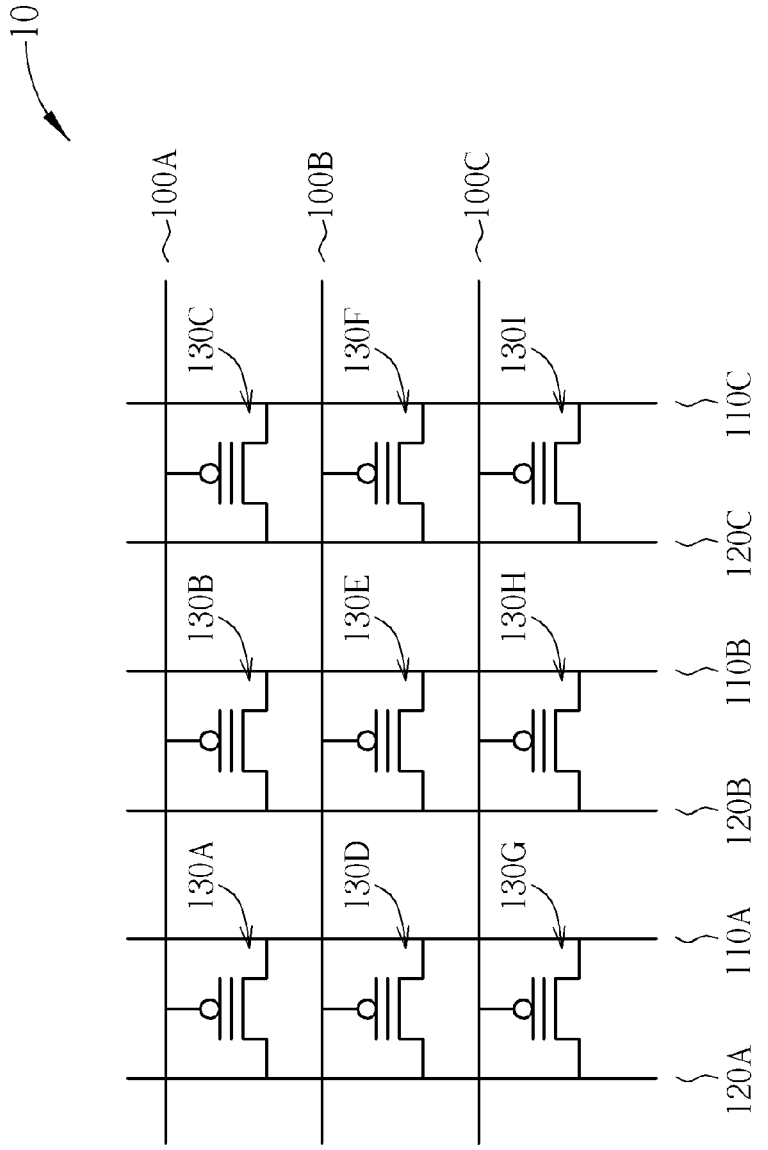
10
100A
100B
100C
130A
130B
130C
130D
130E
130F
130G
130H
130I
120A
110A
120B
110B
120C
110C

FIG. 1

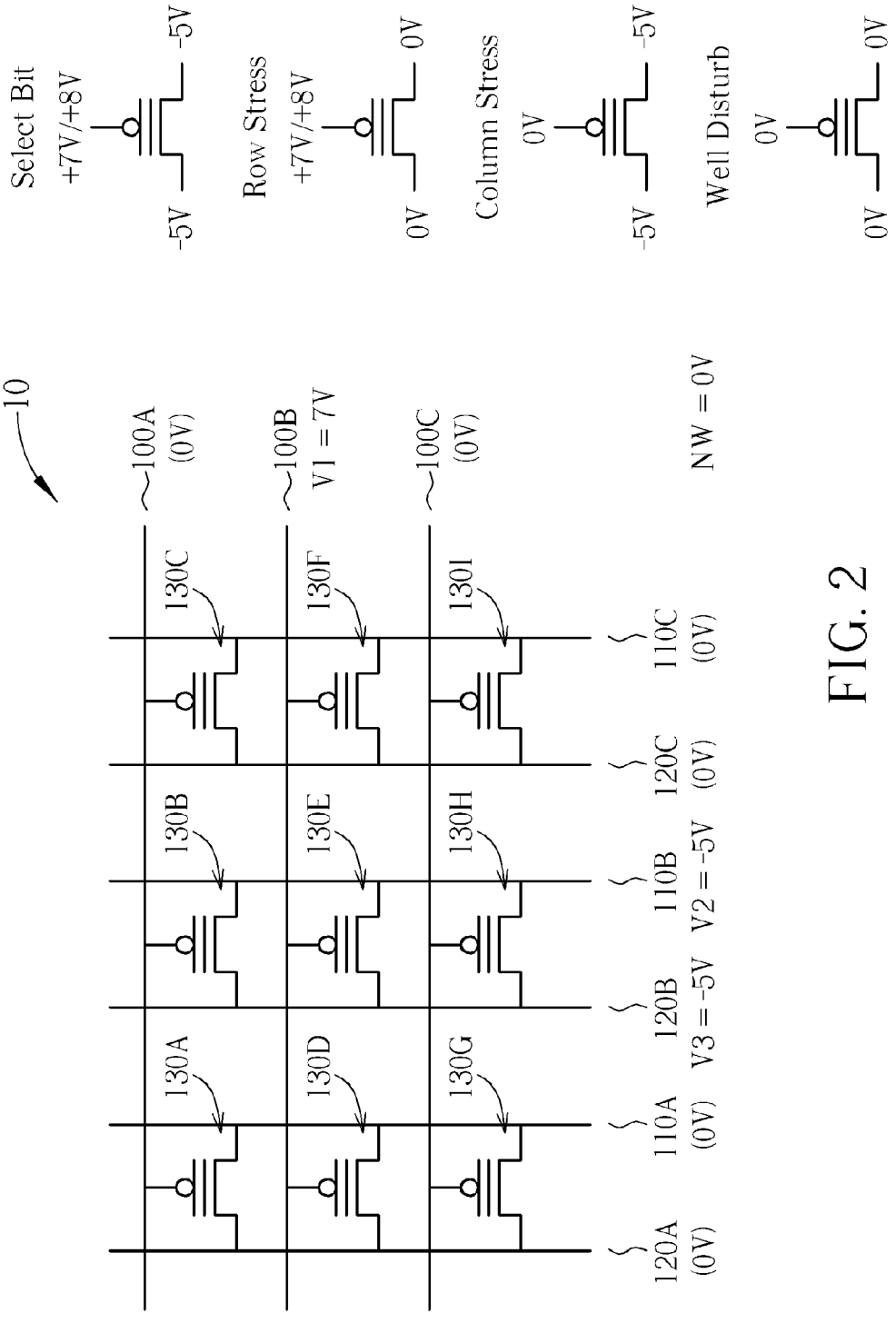
Select Bit
+7V/+8V
-5V
-5V
Row Stress
+7V/+8V
0V
0V
Column Stress
0V
-5V
-5V
Well Disturb
0V
0V
0V
10
100A
(0V)
100B
V1 = 7V
100C
(0V)
130A
130B
130C
130D
130E
130F
130G
130H
130I
120A
(0V)
110A
(0V)
120B
V3 = -5V
110B
V2 = -5V
120C
(0V)
110C
(0V)
NW = 0V

FIG. 2

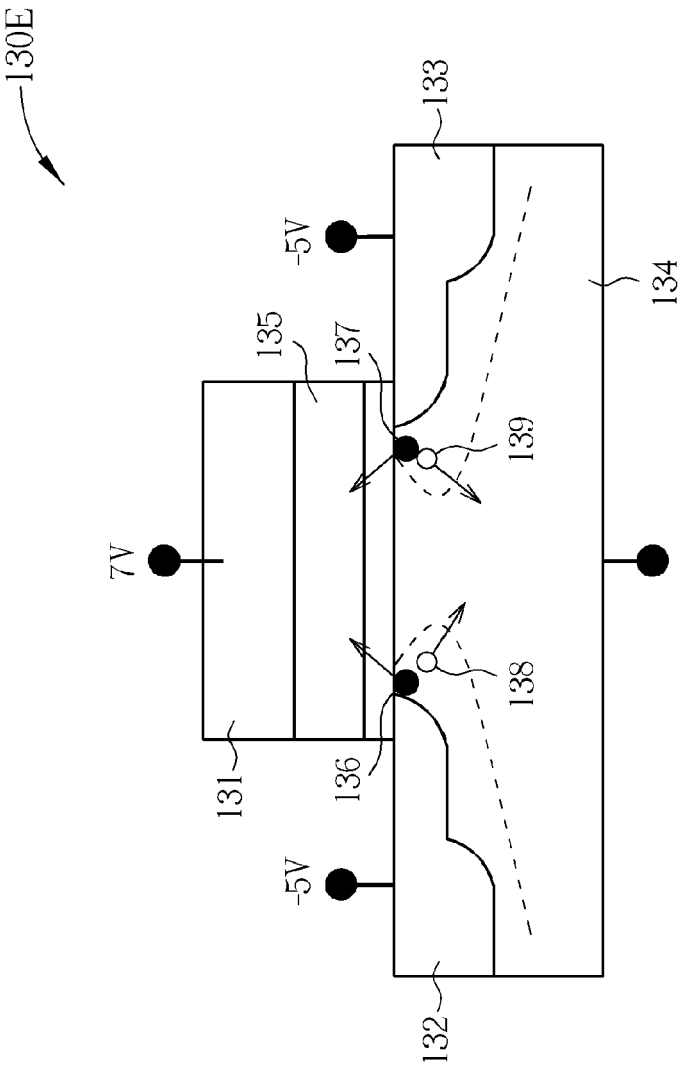

FIG. 3

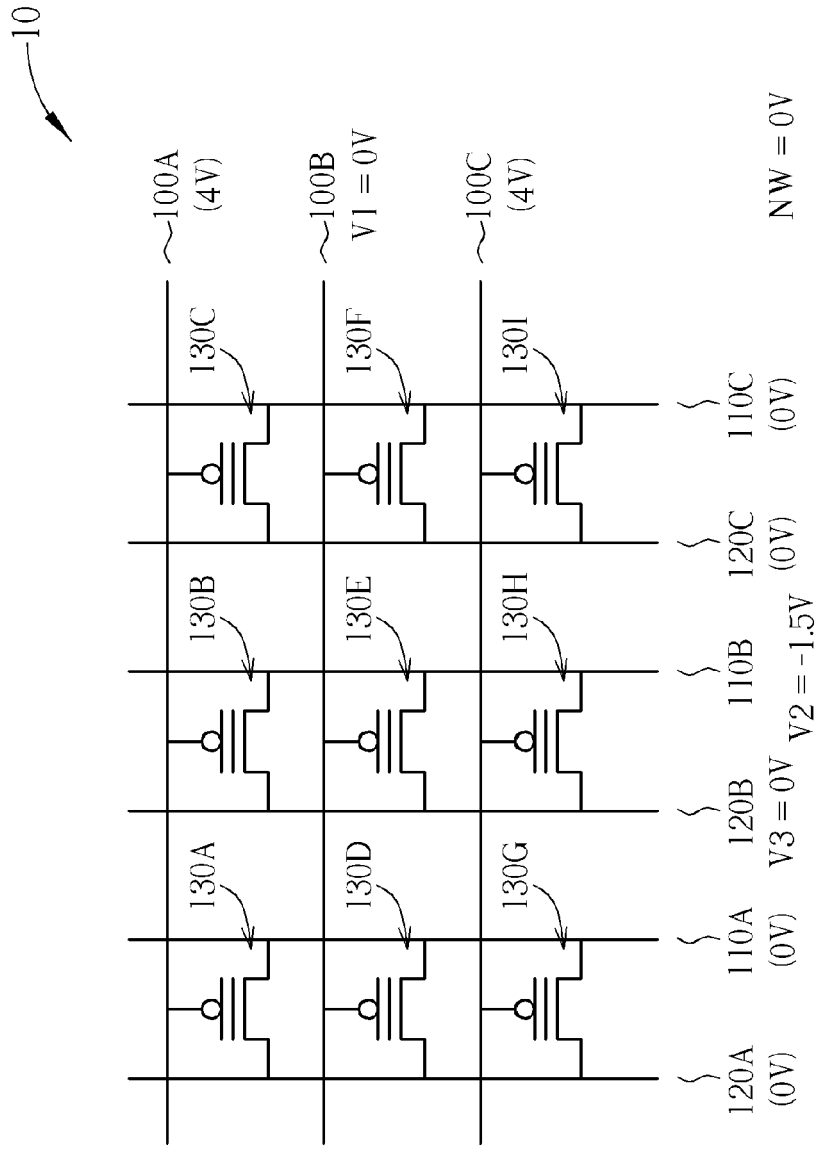
10
100A
(4V)
100B
V1 = 0V
100C
(4V)
NW = 0V
130A
130B
130C
130D
130E
130F
130G
130H
130I
120A
(0V)
110A
(0V)
120B
V3 = 0V
110B
V2 = -1.5V
120C
(0V)
110C
(0V)

FIG. 4

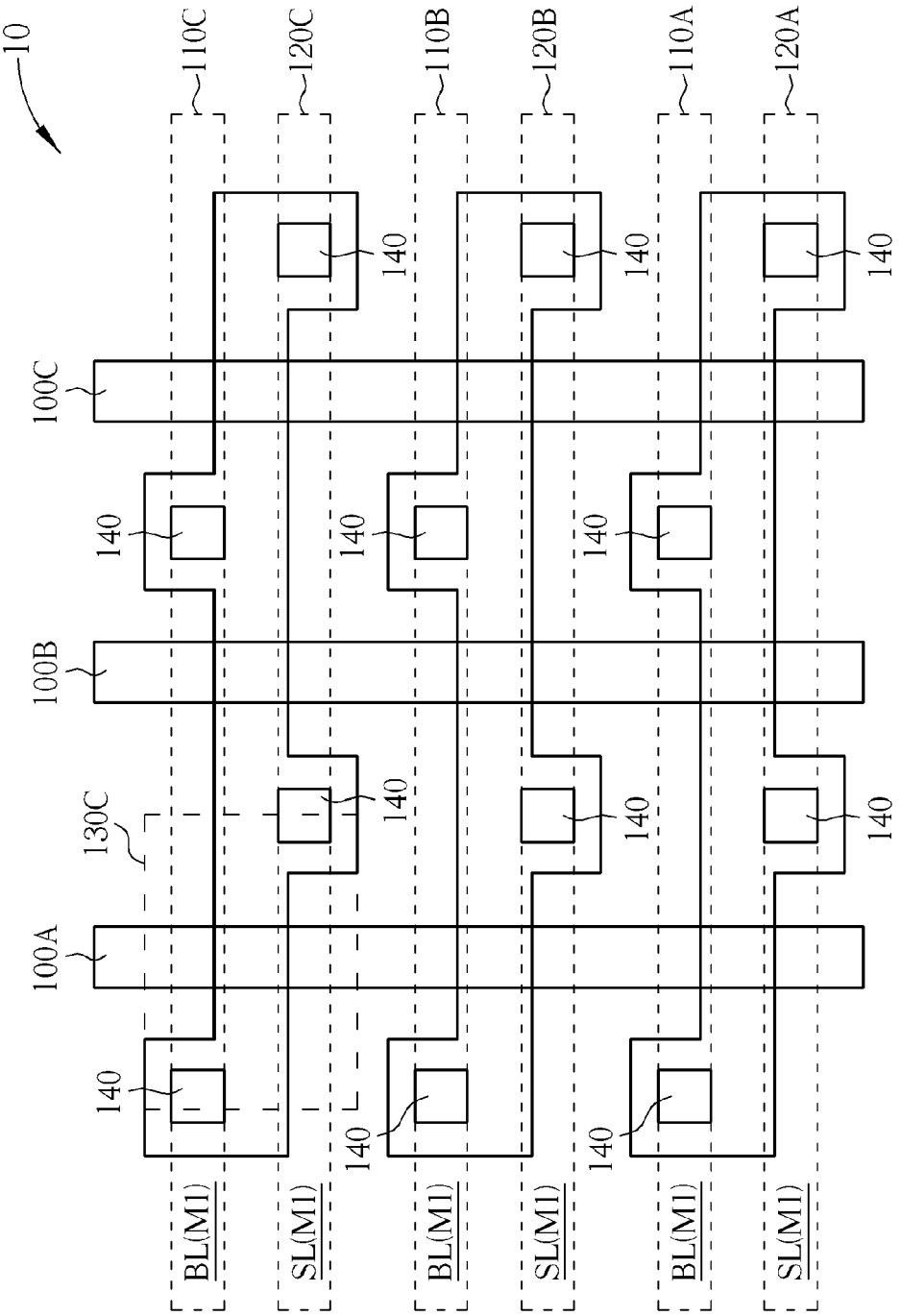
10
110C
120C
110B
120B
110A
120A
100C
100B
100A
130C
140
BL(M1)
SL(M1)

FIG. 5